Metabarcoding and Metagenomics 9: 297–313 (2025)
DOI: 10.3897/mbmg.9.152554

Research Article

Mind the gap: DNA barcoding on *Pachygrapsus* crabs

Cátia Alves[1,2,3,4,5,6], Lara Baptista[1,2,3,5,7], Patrícia Madeira[1,2,3,5], Nuno Vaz Álvaro[8], António Santos[2,4,6], Harald Meimberg[9], Sérgio P. Ávila[1,2,3,4,5], Manuel Curto[2,6]

1 *CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Pólo dos Açores, 9501-801 Ponta Delgada, Portugal*
2 *BIOPOLIS Program in Genomics, Biodiversity and Land Planning, CIBIO, Campus de Vairão, 4485-661 Vairão, Portugal*
3 *UNESCO Chair – Land Within Sea: Biodiversity & Sustainability in Atlantic Islands, Universidade dos Açores, R. Mãe de Deus 13A, 9500-321 Ponta Delgada, Portugal*
4 *Faculdade de Ciências da Universidade do Porto, Rua do Campo Alegre, 1021/1055, 4169-007 Porto, Portugal*
5 *MPB - Marine Palaeontology and Biogeography Lab, Universidade dos Açores, Rua da Mãe de Deus, 9501-801 Ponta Delgada, Portugal*
6 *CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Universidade do Porto, Campus de Vairão, Rua Padre Armando Quintas n° 7, 4485-661 Vairão, Portugal*
7 *Royal Netherlands Institute for Sea Research, Landsdiep 4, 1797 SZ 't Hoorntje, Texel, Netherlands*
8 *Universidade dos Açores, Instituto de Investigação e Tecnologias Agrárias e do Ambiente Campus de Angra do Heroísmo Rua Capitão João d'Ávlia, Pico da Urze PT, 9700-042 Angra do Heroísmo, Portugal*
9 *Department of Ecosystem Management, Climate and Biodiversity, Institute of Integrative Nature Conservation Research, BOKU University, Vienna, Austria*
Corresponding author: Cátia Alves (catia.alves@cibio.up.pt)

Abstract

DNA barcode databases for marine invertebrates faces significant challenges, with several species revealing taxonomic issues like misidentification, ambiguity or shared barcodes. This highlights the importance of the availability of robust genetic data to resolve taxonomic uncertainties and improve species identification, which in many groups of invertebrates, like crustaceans, is lacking. Based on molecular data from the Cytochrome *c* oxidase I 5' gene region (*COI*) for barcoding, this study highlights some of these challenges by addressing existing taxonomic ambiguities and highlighting the scarcity of genetic data for several of the shore crab genus *Pachygrapsus* Randall, 1840. As expected, molecular data availability is highly skewed towards taxa more prevalent in the Northern Hemisphere. Our analysis reveals a clear barcoding gap amongst *Pachygrapsus* species, with average interspecific *p*-distances of 18.75%, exceeding previously reported values for crustaceans. Furthermore, high intraspecific variation (0.70% average) suggests potential taxonomic issues, particularly within *P. gracilis* (de Saussure, 1857). East and west Atlantic *P. gracilis* populations exhibit substantial divergence (3.6% *p*-distance), raising the possibility of ongoing speciation due to limited gene flow across the Atlantic Ocean. Similarly, populations of *P. crassipes* in the east and west North Pacific exhibit substantial divergence (3.3% *p*-distance), which suggests possibly another speciation process. *COI* barcoding proves effective for species identification within *Pachygrapsus* and in unravelling intraspecific patterns indicative of cryptic diversity guiding further taxonomic revisions integrating morphological, genetic and ecological data.

Key words: Barcode, COI, Crustacea, cryptic species, haplotype network, phylogeny

OPEN ACCESS

Academic editor: Baruch Rinkevich
Received: 10 March 2025
Accepted: 17 May 2025
Published: 29 July 2025

Citation: Alves C, Baptista L, Madeira P, Vaz Álvaro N, Santos A, Meimberg H, Ávila SP, Curto M (2025) Mind the gap: DNA barcoding on *Pachygrapsus* crabs. Metabarcoding and Metagenomics 9: e152554. https://doi.org/10.3897/mbmg.9.152554

Copyright: © Cátia Alves et al.
This is an open access article distributed under terms of the Creative Commons Attribution License (Attribution 4.0 International – CC BY 4.0).

Introduction

Accurate species identification and the production of checklists validated by experts are fundamental to biodiversity and biogeography studies, conservation efforts and ecological monitoring (Bresler et al. 1999; Bedulli et al. 2002; Ávila et al. 2019). The morphology-based system used for identifying species in a traditional taxonomic approach is challenging given the current taxonomic gap, i.e. the trend for reduction in the number of expert taxonomists (Hebert et al. 2003; Ebach and Holdrege 2005), the discovery of cryptic diversity (Hebert et al. 2004) and accurate species identification of larval and juvenile forms (Kroh et al. 2012; Keskin and Atar 2013). To solve these issues, researchers started combining traditional approaches with molecular data, such as DNA barcoding, a method that mostly relies on the sequence of the 5' end of the mitochondrial Cytochrome *c* oxidase I gene (*COI*) to provide unique molecular identifiers – "barcode" – specific for each species (Hebert et al. 2003). The single-use of mitochondrial markers has been a subject of ongoing debate and research (Eberle et al. 2020), with studies that support the barcoding initiative, yet acknowledging the need for adjustment of the DNA region and length to produce an accurate representation of the total genetic variability in the gene, leading to proper species delimitation (Roe and Sperling 2007). The standardisation of the usage of a specific genomic region as a barcode led to a coordinated effort between researchers, allowing the creation of homologous sequence data, facilitating the construction of more comprehensive genetic libraries and complex studies (Caterino et al. 2000). It can be used as powerful tool for the evaluation of broader taxonomic and geographical evolutionary patterns (Munch et al. 2008) or even the investigation of intraspecific patterns (Hajibabaei et al. 2007). Furthermore, barcode libraries play a crucial role in the building of custom reference databases for metabarcoding studies providing, for example, full use of non-destructive/non-lethal sampling protocols, such as those used in environmental DNA to describe species habits and biological communities (Taberlet et al. 2012; Djurhuus et al. 2020).

The ability of DNA barcoding to reliably distinguish between closely-related species depends on the existence of a "Barcoding Gap (BG)", which can be observed whenever the divergence amongst conspecific organisms is smaller than that amongst organisms from different species (Puillandre et al. 2012). In the GenBank and BOLD databases, several barcode markers – *COI*, MatK, *RbcL*, RbcLa, trnH-psbA, ITS, ITS2 – make up reference libraries used in metabarcoding studies. During the 8th International Barcode of Life Conference (Trondheim, Norway, 2019), a sub-sample of major taxonomic groups (crustaceans, echinoderms, molluscs and polychaetes) from BOLD data was curated and revealed significant issues in the barcoding coverage of marine invertebrates, with 39% of species reviewed having problems like misidentification, ambiguity or shared barcodes (Radulovici et al. 2021). According to a gap-analysis of the Iberian macroinvertebrate community (especially Crustacea, Mollusca and Polychaeta), around 63% of the species lacked a DNA barcode and a relevant proportion of species were flagged for significant intraspecific divergence and possible hidden diversity (Leite et al. 2020).

In 2005, Poupin et al. (2005) revised the genus *Pachygrapsus* Randall, 1840 (Arthropoda, Crustacea) and acknowledged 12 species, but suggested that

P. propinquus De Man, 1908 should be a consensually accepted junior synonym of *P. minutus* A. Milne-Edwards, 1873. This remained until March 2025, when *P. propinquus* was considered as *Metopograpsus thukuhar* (Owen, 1839), which is another entirely different species, by Sammy De Grave (WoRMS taxonomic editor). Additionally, in 2005, Schubart et al. (2005) recovered the species *P. socius* Stimpson, 1871 for all Pacific representatives of *P. transversus* (Gibbes, 1850), given morphological and genetic differences detected amongst the populations. Today, a total of 13 species are accepted and listed in the WoRMS database (WoRMS 2025). The genus *Pachygrapsus* is dispersed through intertidal zones at tropical and temperate latitudes (Fig. 1), being found amongst rocks, algae, rubble and mangroves (Poore and Ahyong 2023).

In the present study, we evaluated the possible presence of barcoding and databases gaps in the crustacean genus *Pachygrapsus*. Considering the importance of complete barcode libraries and the ecological role of marine crustaceans in macroinvertebrate communities, we assessed the available barcodes and genetic data of *Pachygrapsus* specimens and contributed a new DNA barcode sequence for the species *P. maurus* (Lucas, 1846) to BOLD. Furthermore, we explored whether the existing database is informative to undercover wider phylogeographic patterns across species.

Methods

Specimen collection

Eight samples of *P. maurus* were collected from the intertidal zone at two islands of the Azores Archipelago: three samples of one pereopod per individual and four samples of full adult crabs, collected in Negrito (Terceira Island) in September 2021 and February 2024, respectively; one pereopod sample from Lajes do Pico (Pico Island) was collected in August 2021. The crabs were hand-caught during low tides. For the collection of pereopods, some pressure was applied and the crab itself released the member through autotomy. The samples were preserved in ethanol 96% until DNA extraction.

DNA Extraction, Amplification and Sequencing

DNA extraction from the pereopod muscle tissue was performed using the commercial kits EasySpin® Genomic DNA Tissue (Citomed, Lisbon, Portugal) for *P. maurus* samples of 2021 and EZ-10 Spin Column Plasmid DNA Miniprep Kit (Bio Basic Inc.) for *P. maurus* samples of 2024. DNA quality and integrity were evaluated with 0.8% agarose gel electrophoresis.

Fragments of the mitochondrial gene Cytochrome *c* oxidase subunit I (*COI*) were amplified by PCR, performed in 10 µl volumes, containing 1 µl of DNA, 5 µl of Master Mix (Qiagen), 1 µl of each primer (10 µM) and 2 µl H_2O. Primers jgLCO1490 / jgHCO2198 (Geller et al. 2013) were used with the following cycling profile: 95 °C for 15 min; 35 cycles of 95 °C for 30 s, 52 °C for 30 s, 72 °C for 30 s; 72 °C for 10 min. The efficiency of the amplification reactions and the quality of the DNA products were assessed with electrophoretic runs in 2% (w/v) agarose gel with GelRed (DNA fluorescent dye, BioTarget™). PCR product purification was performed via ExoSAP cleanup 10 µl volumes, entailing 1 µl of DNA, 1 µl of TRR Buffer, 1 µl of TRR, 0.5 µl

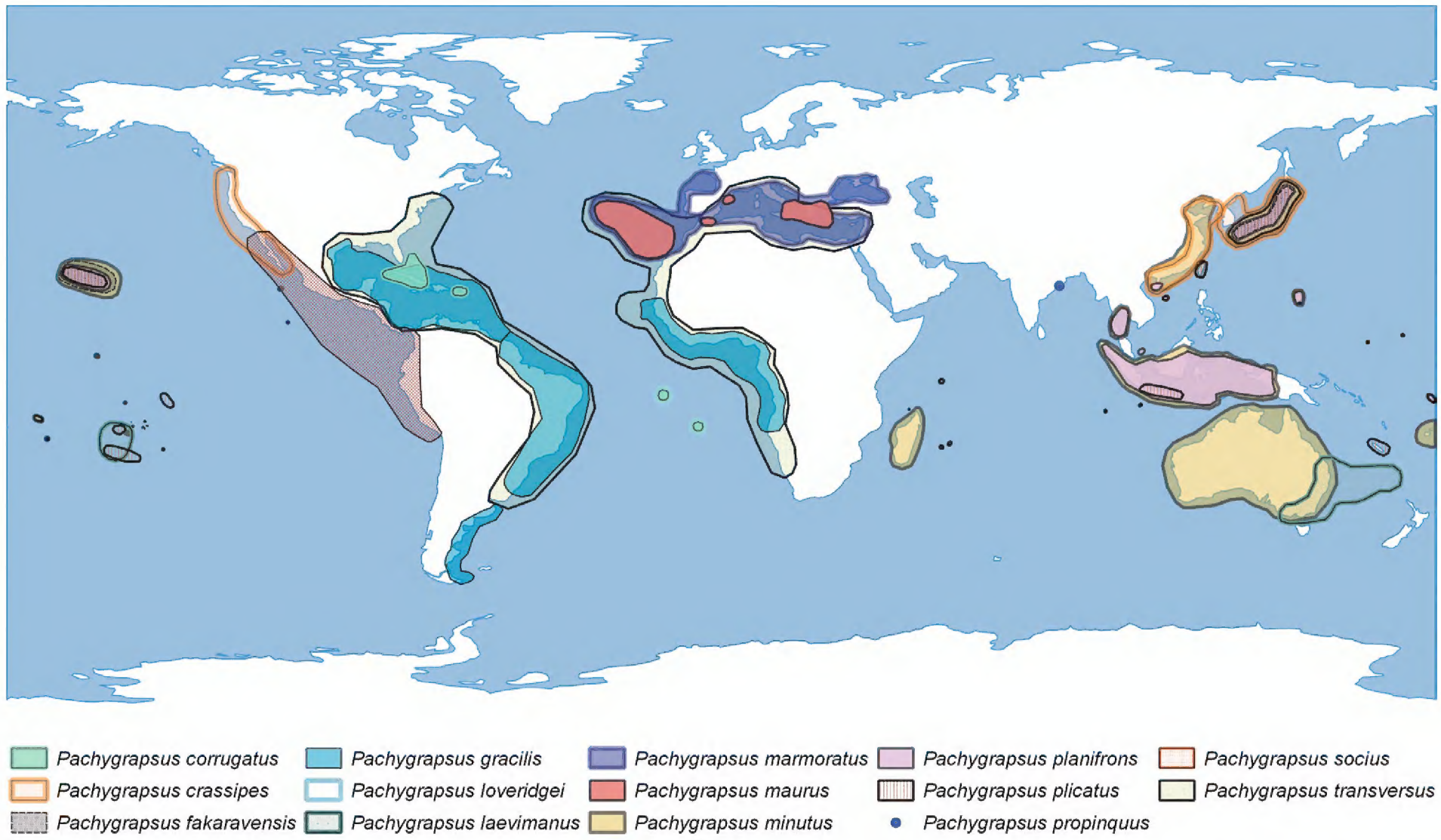

Figure 1. Geographical distribution of all species of the *Pachygrapsus* genus (modified from Poupin et al. (2005)). This map was created using the open-source software QGIS 3.10.1 (http://www.qgis.org/).

of Primer and 7 µl of H_2O. The cycling conditions were as follows: 37 °C for 15 min and 87 °C for 15 min. Finally, bi-directional Sanger sequencing with amplification primers was conducted at CTM – Centre for Molecular Analysis (Vairão, Portugal).

Dataset assemblage

At the time of data collection, all *COI* 5' region sequences (the norm used for DNA-barcoding in metazoa) available for the *Pachygrapsus* genus were retrieved from GenBank (https://www.ncbi.nlm.nih.gov/genbank/) and BOLD (https://boldsystems.org/). This amounted to 332 haplotypes (Table 1) belonging to 10 species, plus eight *COI* sequences from *P. maurus* obtained specifically for this study. New sequences of *P. maurus* were deposited in GenBank under the accession numbers PQ570873-76. The four specimens' representatives for the species barcode are deposited in BOLD under the bin number AGP6456. After inspection, several sequences were found to be shared haplotypes and, when available, meta-information regarding geographical origin was retrieved from supplementary material of the original publications allowing us to reconstruct the original datasets of Cassone and Boulding (2006) for *P. crassipes* (Randall, 1840), and Fratini et al. (2016) for *P. marmoratus* (Fabricius, 1787). The full *Pachygrapsus* dataset of 1,032 sequences is available in Suppl. material 1.

Data analyses

The sequences were edited and aligned using the Clustal Omega algorithm executed on Unipro UGENE (Okonechnikov et al. 2012). Sequences of the mitochondrial coding *COI* gene were translated into amino acids with AliView

Table 1. Dataset of 340 COI 5' region sequences of *Pachygrapsus* species used in this study. The new sequences of *P. maurus* are highlighted in bold.

Taxa	Data Repository	*COI* Accession numbers
P. crassipes Randall, 1840	BOLD	ARIS007-19, BIPP137-17, BIPP149-17, BIPP150-17, BIPP176-17, BIPP182-17, BIPP188-17, DISA118-18, DISA120-18, DISA121-18, DISA562-19, DISA632-19, DISA774-19, DISA881-19, YUN001-16, YUN002-16, ZPC024-13, ZPC033-13, ZPC042-13, ZPC046-13, ZPC056-13, ZPC057-13, ZPC058-13, ZPC060-13, ZPC062-13, ZPC066-13, ZPC072-13, ZPC074-13, ZPC078-13, ZPC079-13, ZPC088-13, ZPC1206-20, ZPC205-14, ZPC211-14, ZPC212-14, ZPC227-14, ZPC291-15, ZPC311-15, ZPC341-15, ZPC342-15, ZPC343-15, ZPC346-15, ZPC404-15, ZPC415-15, ZPC432-15, ZPC458-15
	GenBank	AY951986 - AY952139
	GenBank	NC_021754
	BOLD	ZPC056-13–ZPC058-13
P. fakaravensis Rathbun, 1907	GenBank	MN184703
P. gracilis (de Saussure, 1857)	GenBank	KU163295
	GenBank	KU163296, MF490106, MN183930, MN183941
P. laevimanus Stimpson, 1858	GenBank	MH931286 - MH931291
P. marmoratus (Fabricius, 1787)	GenBank	JF930650 - JF930653
	GenBank	JF930655 - JF930682
	GenBank	JQ305921, KT988331, KY639429, MN185365, MT920155
	GenBank	JQ306087–JQ306092
	GenBank	KX529672–KX529696
	GenBank	KX549320–KX549335
	BOLD	MLALE002-09, MLALE018-09, MLALE020-09, MLALN027-10
***P. maurus* (Lucas, 1846)**	**BOLD**	**AGP6456**
	GenBank	**PQ570873–PQ570876**
P. minutus A. Milne-Edwards, 1873	GenBank	KANBI1398, KANBI1554, KANBI261, KANBI450, MBMIA092, MW277877, MW278021, MW278717, MW278837
P. planifrons De Man, 1888	BOLD	MBMIA059-06, MBMIA093-06
	GenBank	MW277908, MW278022, MW278714, MW278729, MW278770
P. plicatus (H. Milne Edwards, 1837)	BOLD	KANBI298-19, MBMIA042-06
	BOLD	MBMIA575-06–MBMIA577-06
	GenBank	MW277904
P. transversus (Gibbes, 1850)	BOLD	DSCMX007-21, DSCMX010-21, DSCMX038-21, DSCMX041-21
	BOLD	FPMAR177-08
	GenBank	KU313192, MN184059, MN184072, MN184077

(Larsson 2014), for the detection of stop codons and frameshifts in their open reading frames. For phylogenetic analysis, the programme IQ-TREE 2 (Minh et al. 2020) was used to construct a Maximum Likelihood phylogenetic tree from the *COI* alignment to infer the evolutionary history. This was conducted under the TN+F+I+G4 model of molecular evolution, estimated as the most adequate by a built-in model selection tool in IQ-TREE 2 using the Akaike Information Criterion. As outgroup, four sequences of *Grapsus adscensionis* (Osbeck, 1765) from GenBank (Acc. numbers MZ368560 to MZ368563) were added to the analysis. The outgroup species was selected, based on its phylogenetic proximity to the ingroup, as established in previous studies (e.g. Ip et al. (2015)). Branch support was calculated using 1,000 bootstrap replicates (Felsenstein 1985). Tree graphics were formatted using FigTree (Rambaut 2015).

The genetic structure amongst the *Pachygrapsus* species was analysed with TCS 1.22 (Clement et al. 2000) using *COI*, generating a statistical parsimony haplotype network at the 95% connection limit. The tcsBU web-based programme (Santos et al. 2016) was utilised to integrate the genetic structure derived from TCS with the geographical distribution by countries, where the *Pachygrapsus* specimens were collected. Levels of *COI* nucleotide divergences amongst conspecific and congeneric individuals were assessed by calculation of pairwise distances using both the *p*-distances and the Kimura 2-Parameters model (K2P), with standard error estimated by a bootstrap procedure (1,000 replicates), following the approach of Srivathsan and Meier (2012). These nucleotide divergences were obtained using the programme MEGA11 (Tamura et al. 2021). The BG was assessed for each species by calculating the difference between the minimum interspecific distance and the maximum intraspecific distance, amongst a set of congeneric species.

Results

Currently, the *Pachygrapsus* genus includes 14 species that have been unequally researched since the initial description of the genus by Randall (1840). For the species *P. loveridgei* Chace, 1966, which was solely described and stored in museum collections, genetic data are not available for further studies and to confirm the species status. Genetic data availability also appears to be dependent on abundance and geographic distribution for other *Pachygrapsus* species. The mitochondrial markers *COI* and gene encoding for the small ribosomal subunit RNA (*16S*) are the most utilised in studies (Table 2). Three species (*P. corrugatus* (von Martens, 1872), *P. maurus* and *P. socius*) did not contain any information for the barcoding region (5' end of *COI*). Of note is *P. socius*, that, despite having sequences available for *COI* (22 sequences from the Jerry-Pat region), none covered the established 'Folmer' barcode region. Based on sequence data availability alone, the species *P. crassipes*, *P. marmoratus* and *P. transversus* are the most studied species of the genus, with a broad range of research topics (e.g. ecology, genetics, physiology).

In total, 340 unique sequences from 10 ingroups and one outgroup species were analysed and results are shown in the Maximum Likelihood tree (Fig. 2; extended tree at Suppl. material 2). Members of the same species formed monophyletic groups with high support values (SH-aLRT ≥ 90%, UFBoot ≥ 98%). The pairs of species that formed clades with high support were *P. crassipes* + *P. transversus* (SH-aLRT: 94%, UFBoot: 88%) and *P. marmoratus* + *P. maurus* (SH-aLRT: 99.6%, UFBoot: 100%).

The TCS haplotype network displays the genetic relationships amongst the *Pachygrapsus* species (Fig. 3), inferred from sequences of the mitochondrial *COI* barcoding region. The network reveals distinct clusters corresponding to species delimitation, suggesting clear genetic differentiation between all of them. A genetic divergence pattern amongst west and east Atlantic populations of *P. gracilis* (de Saussure, 1857) was observed forming two small and separated clusters. *P. crassipes* also revealed high intraspecific variance with two major clusters – corresponding to USA (east North Pacific) and South Korea (west North Pacific, Sea of Japan) populations (Fig. 3) – separated by six mutational steps between the closest haplotypes and with some derivative haplotypes.

Table 2. Summary of the genetic data available in GenBank for the species of the genus *Pachygrapsus* (data collection in June 2024).

	Genetic marker	*Pachygrapsus* species												Total
		P. corrugatus	*P. crassipes*	*P. fakaravensis*	*P. gracilis*	*P. laevimanus*	*P. marmoratus*	*P. maurus*	*P. minutus*	*P. planifrons*	*P. plicatus*	*P. socius*	*P. transversus*	
Complete cds/genome	Arginine kinase, mRNA						1							1
	Heat shock protein, mRNA						4							4
	Hyperglycemic hormone CHH A OX						1							1
	Hyperglycemic hormone CHH B PO						2							2
	Hyperglycemic hormone precursor						1							1
	Mitochondrial genome		2	1			2							5
	Na-K-ATPase alpha-isoform subunit C						1							1
	NA-K-ATPase alpha-isoform subunit D						1							1
Partial sequence	*COI* Folmer barcoding region		155	1	6	6	88		10	7	6		9	288
	COI Jerry-Pat region				48							22	66	136
	Enolase		1	1	1	1	1	1	1		2	1	1	9
	Flanking regions of mariner-like transposon						22							22
	Heat shock protein, mRNA						7						1	7
	Histone (H3)		1	1	1	1	2	1	1	1	2	1	2	11
	LSU, rRNA		4				3							7
	L-lactate dehydrogenase												1	0
	Na-K-ATPase subunit alpha						1							1
	GAPDH			1			1					1		2
	Pgcra spliceosome-associated protein		1	1			1				1			4
	Pgcra ATP synthase beta subunit		1	1			1							3
	Pgcra arginine kinase		2	1	1		1				2	1		7
	Phosphoenolpyruvate carboxykinase						1				2			3
	Ribosomal protein L10												2	0
	Small subunit ribosomal RNA			1										1
	Sodium-potassium adenosine triphosphatase												1	0
	Na-K ATPase alpha subunit		1	1	1	1	1	1	1		2	1	1	9
	Vacuolar ATP synthase subunit B						1							1
	12S		1	1	1	1	2	1	2	2	2	1	2	13
	16S	1	13	2	7	2	9	2	2	1	3	13	27	41
	28S						1							1
	Microsatellites						25							25
	Total	1	182	13	66	12	181	6	17	11	22	41	113	

The intraspecific *COI* divergences for *P. fakaravensis* Rathbun, 1907 could not be assessed given that a single individual composes the entire sample pool in this study. The remaining taxa had low intraspecific divergences (Table 3), apart from *P. gracilis* (average 3.6%) and *P. crassipes* (maximum 3.3%). Congeneric divergences varied from 11.3% to 19.6% as comparisons were established between *P. marmoratus* - *P. maurus* and *P. planifrons* - *P. transversus*, respectively. In all *Pachygrapsus* species – except *P. fakaravensis*, for which calculation was not possible due to low sample size – a barcoding gap was detected (BG > 0) and, since the average between-species sequence divergence (18.75%) is higher than intraspecific distances (0.70%), it enables clear molecular species identification (Table 3).

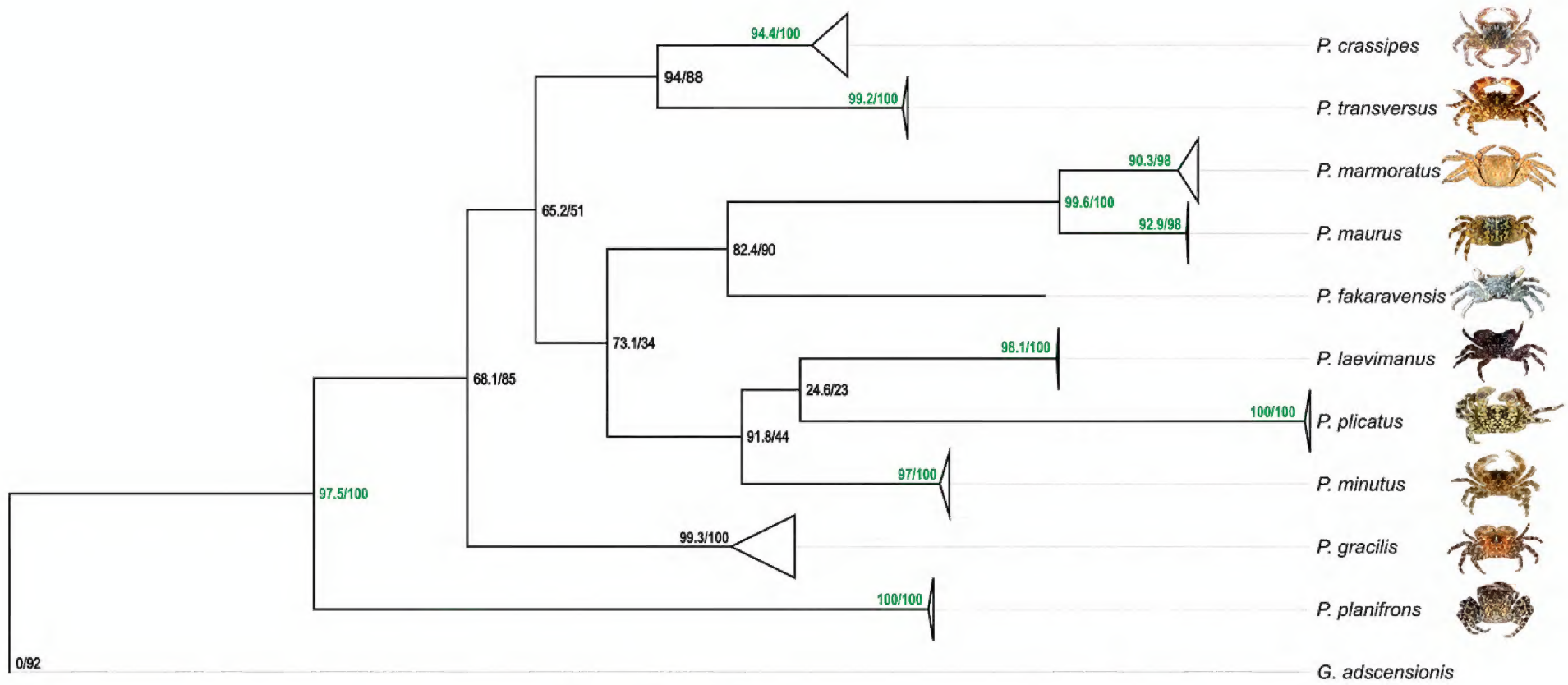

Figure 2. Maximum Likelihood tree depicting genetic distance amongst *Pachygrapsus* species, based on the 340 unique haplotypes from the barcoding region of the *COI* marker (collapsed sequences per species: 201 *P. crassipes*, 9 *P. transversus*, 88 *P. marmoratus*, 8 *P. marmoratus*, 8 *P. maurus*, 1 *P. fakaravensis*, 6 *P. laevimanus*, 6 *P. plicatus*, 9 *P. minutus*, 5 *P. gracilis* and 7 *P. planifrons*). The selected model of evolution was TN+F+I+G4 for 10 *Pachygrapsus* species and four outgroup taxa. Log-likelihood of consensus tree: -5598.470, proportion of invariable sites 0.536, gamma shape alpha: 0.789. Node labels refer to SH-aLRT support (%) / ultrafast bootstrap support (%). Nodes with both SH-aLTR and UF support values above 90% have the respective values highlighted in green. Picture representative of the *Pachygrapsus* species included in the phylogenetic reconstructed, not to scale. Phylogenetic tree inferred with IQ-TREE 2 (Minh et al. 2020).

Table 3. Pairwise mitochondrial *COI* nucleotide uncorrected p-distances for calculation of the BG in the genus *Pachygrapsus*. BG was calculated subtracting the maximum intraspecific distance from the minimum congeneric distance.

Taxon	Divergence	Min distance	Mean distance	Max distance	BG
P. marmoratus	Intraspecific	0.000	0.005 ± 3.83E-05	0.017	0.096
	Congeneric	0.113	0.190 ± 0.010	0.236	
P. maurus	Intraspecific	0.000	0.002 ± 2.17E-04	0.003	0.109
	Congeneric	0.113	0.188 ± 0.010	0.222	
P. laevimanus	Intraspecific	0.000	0.001 ± 2.18E-04	0.002	0.144
	Congeneric	0.146	0.171 ± 0.008	0.203	
P. minutus	Intraspecific	0.000	0.003 ± 0.001	0.014	0.132
	Congeneric	0.146	0.180 ± 0.009	0.224	
P. gracilis	Intraspecific	0.000	0.036 ± 0.009	0.062	0.106
	Congeneric	0.168	0.185 ± 0.009	0.220	
P. planifrons	Intraspecific	0.000	0.003 ± 2.96E-04	0.005	0.191
	Congeneric	0.196	0.198 ± 0.012	0.236	
P. plicatus	Intraspecific	0.000	0.003 ± 0.001	0.005	0.172
	Congeneric	0.177	0.195 ± 0.011	0.222	
P. transversus	Intraspecific	0.000	0.002 ± 2.65E-04	0.005	0.151
	Congeneric	0.156	0.186 ± 0.010	0.234	
P. fakaravensis	Intraspecific	–	–	–	–
	Congeneric	0.187	0.191 ± 0.010	0.225	
P. crassipes	Intraspecific	0.000	0.009 ± 4.871E-05	0.033	0.121
	Congeneric	0.154	0.190 ± 0.010	0.220	

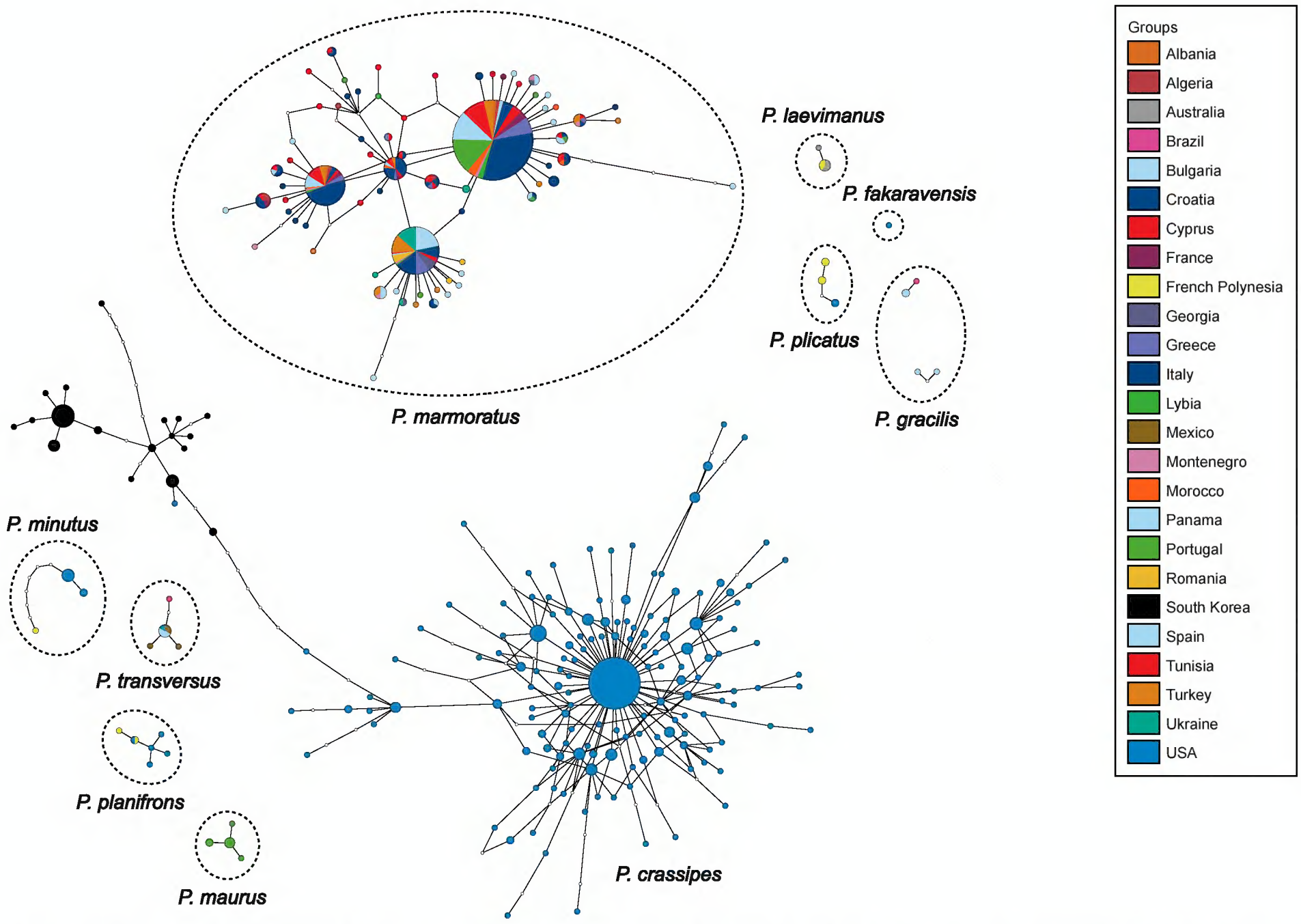

Figure 3. Haplotype network, at 95% parsimony connection limit, for 1,032 COI sequences of individuals from 10 species of the *Pachygrapsus* genus labelled, based on the geographic location of the sampling sites. Each location is represented by distinct colours and each cluster corresponds to a different species. Circle size is proportional to haplotype frequency.

Discussion

The sequences from 10 species of the *Pachygrapsus* genus examined in this study are contained in the *COI* 5' region, usually addressed as the barcoding region. Despite recent efforts to create a baseline of barcode libraries by projects such as the International Barcode of Life Consortium, many species still lack genetic information available. An example of this in the genus *Pachygrapsus* is the lack of *COI* sequences available for the species *P. corrugatus* and *P. loveridgei.* In the case of *P. socius*, the *COI* sequences available are from the so-called Jerry-Pat marker region, which, despite its informative value, constitutes a different region of *COI* (3' direction) and is not comparable with the remaining species in this study. Even though most of the science community uses the barcoding region, some indicate that this region is no better than other regions downstream in *COI* and recommend maximising the sequence length to optimise the chance of sampling regions of high phylogenetic informativeness and minimise stochastic variation when estimating total divergence (Roe and Sperling 2007). The use of the full-length *COI* gene might be facilitated with the establishment of long-read sequencing as the go-to sequencing technology for the integration of both regions (Jeong et al. 2021). In Table 2, the bias in genetic information availability between the *Pachygrapsus* species is evident, with almost no data on species with no commercial interest in underdeveloped countries (Fig. 1). At the other end of the

spectrum lies the crab species, such as *P. crassipes* and *P. marmoratus*, which stand out in the number of studies done since they are highly abundant species (Fusi et al. 2016; Lord and Barry 2017) and inhabit the west coast of the United States of America and the coasts of European countries, respectively. These characteristics tend to influence the research effort dedicated to these species. Even with citizen-science initiatives, these two species also stand out amongst the other *Pachygrapsus* in the number of observations: their reported observations in iNaturalist take on 68.2% and 22.7%, respectively, from the registered observations of this genus (data collected on 2/12/2024 at https://www.inaturalist.org/).

The shore crab genus *Pachygrapsus* was assembled, based on a set of selected morphological diagnosing characters (Poupin et al. 2005), which may not reflect true evolutionary histories. Schubart (2011) was amongst the first to suggest polyphyly in *Pachygrapsus*, based on a molecular phylogeny of the family Grapsidae using the *16S* rRNA gene. Additionally, a later study with a higher phylogenetic resolution from Ip et al. (2015) confirmed that the *Pachygrapsus* genus may not form a natural group, as species were dispersed throughout the Grapsidae family tree, based on both mitochondrial (*16S* rRNA and *12S* rRNA genes) and nuclear markers (sodium-potassium adenosine triphosphatase α-subunit, enolase and histone 3). In our Maximum Likelihood phylogenetic tree (Fig. 2), the relationships between *P. marmoratus* – *P. maurus* and *P. crassipes* – *P. transversus* were congruent with the species delimitation of the genus *Pachygrapsus* in literature. Nevertheless, it revealed clades with varying levels of statistical support (SH-aLRT/UFBoot values). Some of the low values may be a result of the polyphyletic nature of this genus (Ip et al. 2015). Phylogenies constructed through the *COI* gene can be very useful, despite some already known caveats. Being a mitochondrial gene, it only shows relationships reflecting the maternal lineage (Zhang and Hewitt 1996). Furthermore, its high mutation rate may increase the possibility of homoplasy that compromises the resolution of deeper relationships, as well as increased possibility of long-branch attraction, especially when some taxa are missing (Phillips and Penny 2003). Thus, the phylogenetic relationships of this genus should consider multi-locus approaches and include information from all species from the ingroup, including from the remaining related genera. Future studies regarding a taxonomic revision of *Pachygrapsus* could reduce taxonomic problems, considering morphologic, genetic (a combination of mitochondrial and nuclear genes) and ecological traits data, to suggest new genus placements for some of its species.

In our results, *P. gracilis* populations from east and west Atlantic had a complete separation of the networks at 95% connection limit (Fig. 3), suggesting that gene flow across the Atlantic is very limited and that it may be a case of cryptic speciation (Hart and Sunday 2007). However, given the small sample used for this study, further sampling will be necessary to ascertain this hypothesis. The geographical distribution of *P. gracilis* in the east Atlantic is described as restricted from Senegal south to Angola (Poupin et al. 2005), despite two records on European waters. The first record happened in hull samples at Bremerhaven in Germany (Gollasch 1999) and the second record in 2013 from the intertidal of a marina in Galician waters (Spain; Cuesta et al. (2016)). One of the clusters of *P. gracilis* (Fig. 3) corresponds to the sequences obtained of the Galician population, which represents an introduction of a non-indigenous species. Given that individuals of *Pachygrapsus* have already been reported on ship's hulls (Wolff 1954) and the *16S* gene was an identical match between individuals from Galicia and Ivory

Coast, it is thought that *P. gracilis* have reached Galician waters by ship hull fouling of tuna vessels originating from the Ivory Coast (for details see: Cuesta et al. (2016)). Additionally, Thiercelin (2016) concluded that *P. gracilis* populations from the northern and southern tropical west Atlantic showed reduced differentiation; hence, no clear geographic pattern existed for the species in this region.

All other *Pachygrapsus* species revealed single clusters at 95% connection limit in the haplotype network analysis (Fig. 3). Of notice was *P. crassipes* which showed a clear differentiation pattern (six mutational steps) between eastern Pacific and Korean populations. Cassone and Boulding (2006) detected highly significant levels of genetic structuring between these populations, concluding that trans-Pacific gene flow has been restricted for quite some time and suggesting their separation as two different species. Even though populations with large effective sizes tend to accumulate more mutations and have a higher number of mutations, for this study, it is not possible to establish such a connection since the size of datasets by species was highly disproportional.

Despite the high sampling and geographical coverage of the sequences obtained for *P. marmoratus*, no apparent intraspecific divergence patterns were observed. This is a species that forms large populations and its resilience – either thermal or salinity tolerance (Jayasundara et al. 2007; Fusi et al. 2016) – allows it to colonise a wide range of habitats, which, possibly combined with other life-history traits, originates the panmixia observed in the haplotypes network (Fig. 3).

In DNA barcoding, a *COI*-based system has been demonstrated to be efficient in discerning at species-level resolution for large taxonomic assemblages of animals, such as insects, birds, fishes and crustaceans (Costa et al. 2007). To discern between congeneric species, a BG has to occur, which was detected for all the *Pachygrapsus* species in this study with enough genetic data available (excluding *P. fakaravensis*; Table 3). For species identification, based on the *COI* gene, it is not possible to set a specific threshold that works for all taxa. For instance, some suggested in lepidopterans that it is 3%, in Thecostraca and Copepoda between 2.1% and 2.6% and in aquatic oligochaete 10% (Hebert et al. 2003; Vivien et al. 2017; Young et al. 2017). Note that divergence values calculated in this study are *p*-distances, while literature values presented here are K2P distances. In this study, it can be ascertained that the *COI* is informative for the decapod genus *Pachygrapsus*, given the presence of BG. Levels of *p*-distances sequence divergence amongst congeneric species averaged 18.75%, which is higher than the 17.16% already reported for decapods (Costa et al. 2007). The relatively wide range of congeneric distances may suggest varying degrees of divergence amongst species, potentially reflecting differences in evolutionary history or the presence of cryptic lineages. Furthermore, the polyphyletic nature of this genus may inflate some of these values since it may need to be broken into multiple ones. Regarding levels of intraspecific variation, common values in various taxa range from 0.25% to 0.30% (Costa et al. 2007), with decapod, or crustacean values in general being described as around 0.38% and 0.48%, respectively (Raupach et al. 2015). In *Pachygrapsus*, the intraspecific variation was higher than the previously described maximum for crustaceans, with an average of 0.70%. Perhaps the level of barcode variation within *Pachygrapsus* species has been overestimated due to the current taxonomic system established for the genus, overlooking possible new species. Given that K2P is reported to inflate genetic divergence, the fact that we have higher divergence values than

reported literature gives us confidence on the existence of the BG and potential cryptic diversity within the *Pachygrapsus* genus. In the case of *P. gracilis*, with an average intraspecific *p*-distance of 3.6%, it has contributed the most to raising the genus-wide average for this measure. Despite this study having a small sample size for *P. gracilis* specimens, the results align with another study based on the *16S* gene from Mexican and Galician populations, where Cuesta et al. (2016) also noticed high intraspecific variation of 2.9%. These extremely divergent values contribute to the possibility of a future separation of east and west Atlantic populations of *P. gracilis* as two different species.

Conclusions

This study highlights the utility of barcoding for discerning congeneric *Pachygrapsus* species, given the presence of a BG for all analysed species, hence confirming the informativeness of the *COI* gene for species-level identification within this crustacean group. The average interspecific *p*-distances divergence of those observed amongst *Pachygrapsus* species is higher than previously reported averages for decapods (Costa et al. 2007). The higher intraspecific *COI* variation observed, averaging 0.70% and exceeding previously reported maxima for crustaceans (Raupach et al. 2015), merits further investigation. This relevant intraspecific variation may be influenced by the current taxonomic classification of *Pachygrapsus*, potentially overlooking cryptic species. The particularly high intraspecific *p*-distance in *P. gracilis* (3.6%), corroborated by previous *16S* gene analysis (Cuesta et al. 2016) and the complete separation of the parsimony networks at 95% connection limit suggest the possibility of distinct east and west Atlantic lineages as potential cryptic species. This reinforces the need for a thorough taxonomic revision of the genus *Pachygrapsus*, integrating morphological, genetic and ecological data. Furthermore, the observed genetic structuring within *Pachygrapsus*, particularly the trans-Pacific divergence in *P. crassipes*, supports previous findings of restricted gene flow and underscores the potential for speciation within this genus. These common trans-oceanic intraspecific patterns highlight the utility of DNA barcoding in going beyond species identification and strengthens the use of DNA barcodes also for broad scale assessments via metabarcoding. The bias in genetic databases towards certain *Pachygrapsus* species, likely due to economic and geographic factors, emphasises the need for increased research efforts on less-studied species to gain a more comprehensive understanding of its evolutionary history and biodiversity.

Acknowledgements

We thank Sara Fratinni (University of Florence) and Elizabeth Boulding (University of Guelph) for information used in this study. We thank Alexander M. Weigand (Musée National d'Histoire Naturelle Luxembourg) for the suggestion of the scope of this article. SPA acknowledges the project M1.1.A/INFRAEST CIENT/A/001/2021 - Base de Dados da PaleoBiodiversidade da Macaronésia, funded by Direção Regional da Ciência e Tecnologia, Governo Regional dos Açores and his research contract with BIOPOLIS funded by FCT/2023.07418 CEEECIND. We acknowledge the Regional Government of the Azores for the sampling licences: CCIP 26/2020/DRCT and CCIR 29/2021/DRCTD. This work also benefited from FEDER funds,

through the Operational Programme for Competitiveness Factors – COMPETE and from National Funds, through FCT: UIDB/50027/2020 (CIBIO/INBIO) and POCI-01-0145-FEDER-006821 and LA/P/0048/2020 (CIBIO/INBIO), as well as through the Regional Government of the Azores (M1.1.a/005/Funcionamento-C-/2016, CIBIO-A; M3.3.B/ORG.R.C./002/2021; and M3.3.B/ORG.R.C./005/2021).

Additional information

Conflict of interest

The authors have declared that no competing interests exist.

Ethical statement

No ethical statement was reported.

Funding

This work was supported by National Funds through Fundação para a Ciência e a Tecnologia in the scope of the project UID/50027 - Rede de Investigação em Biodiversidade e Biologia Evolutiva; and PhD grant PRT/BD/151537/2021 to CA. Work was also supported by the European Union's Horizon 2020 Research and Innovation Programme under the Grant Agreement Number 857251.

Author contributions

Conceptualization: HM, MC, AS, CA, LB. Data curation: CA, SPPÁ. Formal analysis: LB, CA, MC, HM, AS. Funding acquisition: PM. Investigation: CA, SPPÁ, NVVÁ. Methodology: HM, MC, AS, CA. Project administration: PM. Resources: PM. Validation: LB, HM, AS, CA. Visualization: AS, HM, LB, CA. Writing - original draft: CA. Writing - review and editing: AS, MC, PM, HM, SPPÁ, LB.

Author ORCIDs

Cátia Alves https://orcid.org/0009-0006-4312-3395
Lara Baptista https://orcid.org/0000-0002-4429-9855
Patrícia Madeira https://orcid.org/0000-0003-4356-9769
Nuno Vaz Álvaro https://orcid.org/0000-0002-6967-7170
António Santos https://orcid.org/0000-0002-0417-8813
Harald Meimberg https://orcid.org/0000-0001-6696-2649
Sérgio P. Ávila https://orcid.org/0000-0003-4317-3051
Manuel Curto https://orcid.org/0000-0002-1630-4653

Data availability

All of the data that support the findings of this study are available in the main text or Supplementary Information.

References

Ávila SP, Melo C, Berning B, Sá N, Quartau R, Rijsdijk KF, Ramalho RS, Cordeiro R, De Sá NC, Pimentel A, Baptista L, Medeiros A, Gil A, Johnson ME (2019) Towards a 'Sea-Level Sensitive' dynamic model: Impact of island ontogeny and glacio-eustasy on global patterns of marine island biogeography. Biological Reviews of the Cambridge Philosophical Society 94: 1116–1142. https://doi.org/10.1111/brv.12492

Bedulli D, Bassignani F, Bruschi A (2002) Use of biodiversity hotspots for conservation of Marine Molluscs: A regional approach. Mediterranean Marine Science 3: 113–121. https://doi.org/10.12681/mms.250

Bresler V, Bissinger V, Abelson A, Dizer H, Sturm A, Kratke R, Fishelson L, Hansen P-D (1999) Marine molluscs and fish as biomarkers of pollution stress in littoral regions of the Red Sea, Mediterranean Sea and North Sea. Helgoland Marine Research 53: 219–243. https://doi.org/10.1007/s101520050026

Cassone BJ, Boulding EG (2006) Genetic structure and phylogeography of the lined shore crab, *Pachygrapsus crassipes*, along the northeastern and western Pacific coasts. Marine Biology 149: 213–226. https://doi.org/10.1007/s00227-005-0197-9

Caterino MS, Cho S, Sperling FAH (2000) The Current State Of Insect Molecular Systematics: A Thriving Tower of Babel. Annual Review of Entomology 45: 1–54. https://doi.org/10.1146/annurev.ento.45.1.1

Clement M, Posada D, Crandall KA (2000) TCS: A computer program to estimate gene genealogies. Molecular Ecology 9: 1657–1659. https://doi.org/10.1046/j.1365-294x.2000.01020.x

Costa FO, deWaard JR, Boutillier J, Ratnasingham S, Dooh RT, Hajibabaei M, Hebert PD (2007) Biological identifications through DNA barcodes: The case of the Crustacea. Canadian Journal of Fisheries and Aquatic Sciences 64: 272–295. https://doi.org/10.1139/f07-008

Cuesta JA, Almón B, Pérez-Dieste J, Trigo JE, Bañón R (2016) Role of ships' hull fouling and tropicalization process on European carcinofauna: New records in Galician waters (NW Spain). Biological Invasions 18: 619–630. https://doi.org/10.1007/s10530-015-1034-9

Djurhuus A, Closek CJ, Kelly RP, Pitz KJ, Michisaki RP, Starks HA, Walz KR, Andruszkiewicz EA, Olesin E, Hubbard K, Montes E, Otis D, Muller-Karger FE, Chavez FP, Boehm AB, Breitbart M (2020) Environmental DNA reveals seasonal shifts and potential interactions in a marine community. Nature Communications 11: 254. https://doi.org/10.1038/s41467-019-14105-1

Ebach MC, Holdrege C (2005) More Taxonomy, Not DNA Barcoding. Bioscience 55: 822–824. https://doi.org/10.1641/0006-3568(2005)055[0823:MTNDB]2.0.CO;2

Eberle J, Ahrens D, Mayer C, Niehuis O, Misof B (2020) A Plea for Standardized Nuclear Markers in Metazoan DNA Taxonomy. Trends in Ecology & Evolution 35. https://doi.org/10.1016/j.tree.2019.12.003

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution; International Journal of Organic Evolution 39: 783–791. https://doi.org/10.1111/j.1558-5646.1985.tb00420.x

Fratini S, Ragionieri L, Deli T, Harrer A, Marino IAM, Cannicci S, Zane L, Schubart CD (2016) Unravelling population genetic structure with mitochondrial DNA in a notional panmictic coastal crab species: sample size makes the difference. BMC Evolutionary Biology 16: 150. https://doi.org/10.1186/s12862-016-0720-2

Fusi M, Cannicci S, Daffonchio D, Mostert B, Pörtner H-O, Giomi F (2016) The trade-off between heat tolerance and metabolic cost drives the bimodal life strategy at the air-water interface. Scientific Reports 6: 19158. https://doi.org/10.1038/srep19158

Geller J, Meyer C, Parker M, Hawk H (2013) Redesign of PCR primers for mitochondrial cytochrome c oxidase subunit I for marine invertebrates and application in all-taxa biotic surveys. Molecular Ecology Resources 13: 851–861. https://doi.org/10.1111/1755-0998.12138

Gollasch S (1999) The asian decapod *Hemigrapsus penicillatus* (de Haan, 1835) (Grapsidae, Decapoda) introduced in European waters: Status quo and future perspective. Helgoländer Meeresuntersuchungen 52: 359–366. https://doi.org/10.1007/BF02908909

Hajibabaei M, Singer GAC, Hebert PDN, Hickey DA (2007) DNA barcoding: How it complements taxonomy, molecular phylogenetics and population genetics. Trends in Genetics 23: 167–172. https://doi.org/10.1016/j.tig.2007.02.001

Hart MW, Sunday J (2007) Things fall apart: Biological species form unconnected parsimony networks. Biology Letters 3: 509–512. https://doi.org/10.1098/rsbl.2007.0307

Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. Proceedings. Biological Sciences 270: 313–321. https://doi.org/10.1098/rspb.2002.2218

Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proceedings of the National Academy of Sciences of the United States of America 101: 14812–14817. https://doi.org/10.1073/pnas.0406166101

Ip BHY, Schubart CD, Tsang LM, Chu KH (2015) Phylogeny of the shore crab family Grapsidae (Decapoda: Brachyura: Thoracotremata) based on a multilocus approach. Zoological Journal of the Linnean Society 174: 217–227. https://doi.org/10.1111/zoj.12235

Jayasundara N, Towle DW, Weihrauch D, Spanings-Pierrot C (2007) Gill-specific transcriptional regulation of Na+/K+-ATPase α-subunit in the euryhaline shore crab *Pachygrapsus marmoratus*: Sequence variants and promoter structure. The Journal of Experimental Biology 210: 2070–2081. https://doi.org/10.1242/jeb.004309

Jeong J, Yun K, Mun S, Chung W-H, Choi S-Y, Nam Y, Lim MY, Hong CP, Park C, Ahn YJ, Han K (2021) The effect of taxonomic classification by full-length 16S rRNA sequencing with a synthetic long-read technology. Scientific Reports 11: 1727. https://doi.org/10.1038/s41598-020-80826-9

Keskin E, Atar HH (2013) DNA barcoding commercially important aquatic invertebrates of Turkey. Mitochondrial DNA 24: 440–450. https://doi.org/10.3109/19401736.2012.762576

Kroh A, Madeira P, Haring E (2012) Species distributions: virtual or real – the case of *Arbaciella elegans* (Echinoidea: Arbaciidae). Journal of Zoological Systematics and Evolutionary Research 50: 99–105. https://doi.org/10.1111/j.1439-0469.2011.00636.x

Larsson A (2014) AliView: a fast and lightweight alignment viewer and editor for large datasets | Bioinformatics | Oxford Academic. https://doi.org/10.1093/bioinformatics/btu531 [October 24, 2024]

Leite BR, Vieira PE, Teixeira MAL, Lobo-Arteaga J, Hollatz C, Borges LMS, Duarte S, Troncoso JS, Costa FO (2020) Gap-analysis and annotated reference library for supporting macroinvertebrate metabarcoding in Atlantic Iberia. Regional Studies in Marine Science 36: 101307. https://doi.org/10.1016/j.rsma.2020.101307

Lord JP, Barry JP (2017) Juvenile Mussel and Abalone Predation by the Lined Shore Crab *Pachygrapsus crassipes*. Journal of Shellfish Research 36: 209–213. https://doi.org/10.2983/035.036.0122

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R (2020) IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Molecular Biology and Evolution 37: 1530–1534. https://doi.org/10.1093/molbev/msaa015

Munch K, Boomsma W, Willerslev E, Nielsen R (2008) Fast phylogenetic DNA barcoding. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 363: 3997–4002. https://doi.org/10.1098/rstb.2008.0169

Okonechnikov K, Golosova O, Fursov M, the UGENE team (2012) Unipro UGENE: A unified bioinformatics toolkit. Bioinformatics (Oxford, England) 28: 1166–1167. https://doi.org/10.1093/bioinformatics/bts091

Phillips MJ, Penny D (2003) The root of the mammalian tree inferred from whole mitochondrial genomes. Molecular Phylogenetics and Evolution 28: 171–185. https://doi.org/10.1016/S1055-7903(03)00057-5

Poore GCB, Ahyong ST (2023) 5271 Marine Decapod Crustacea. A Guide to Families and Genera of the World. CSIRO Publishing, Clayton South, Australia, 916 pp. https://doi.org/10.11646/zootaxa.5271.1.12 [January 16, 2025]

Poupin J, Davie PJF, Cexus JC (2005) A revision of the genus *Pachygrapsus* Randall, 1840 (Crustacea: Decapoda: Brachyura, Grapsidae), with special reference to the South-west Pacific species. Zootaxa 1015: 1. https://doi.org/10.11646/zootaxa.1015.1.1

Puillandre N, Lambert A, Brouillet S, Achaz G (2012) ABGD, Automatic Barcode Gap Discovery for primary species delimitation. Molecular Ecology 21: 1864–1877. https://doi.org/10.1111/j.1365-294X.2011.05239.x

Radulovici AE, Vieira PE, Duarte S, Teixeira MAL, Borges LMS, Deagle BE, Majaneva S, Redmond N, Schultz JA, Costa FO (2021) Revision and annotation of DNA barcode records for marine invertebrates: Report of the 8th iBOL conference hackathon. Metabarcoding and Metagenomics 5: e67862. https://doi.org/10.3897/mbmg.5.67862

Rambaut A (2015) FigTree v1.4.4. https://github.com/rambaut/figtree/releases [January 31, 2025]

Randall JW (1840) Catalogue of the Crustacea brought by Thomas Nutall and J.K. Townsend, from the West Coast of North America and the Sandwich Islands, with descriptions of such species as are apparently new, among which are included several species of different localities, previously existing in the collection of the Academy. Journal of the Academy of Natural Sciences at Philadelphia. https://doi.org/10.5962/bhl.title.119921 [February 13, 2025]

Raupach MJ, Barco A, Steinke D, Beermann J, Laakmann S, Mohrbeck I, Neumann H, Kihara TC, Pointner K, Radulovici A, Segelken-Voigt A, Wesse C, Knebelsberger T (2015) The Application of DNA Barcodes for the Identification of Marine Crustaceans from the North Sea and Adjacent Regions. PLoS ONE: 10(9): e0139421 https://doi.org/10.1371/journal.pone.0139421

Roe AD, Sperling FAH (2007) Patterns of evolution of mitochondrial cytochrome *c* oxidase I and II DNA and implications for DNA barcoding. Molecular Phylogenetics and Evolution 44: 325–345. https://doi.org/10.1016/j.ympev.2006.12.005

Santos A, Cabezas MP, Tavares AI, Xavier R, Branco M (2016) tcsBU: A tool to extend TCS network layout and visualization. Bioinformatics (Oxford, England) 32: 627–628. https://doi.org/10.1093/bioinformatics/btv636

Schubart CD (2011) Reconstruction of phylogenetic relationships within Grapsidae (Crustacea: Brachyura) and comparison of trans-isthmian versus amphi-atlantic gene flow based on mtDNA. Zoologischer Anzeiger 250: 472–478. https://doi.org/10.1016/j.jcz.2011.06.003

Schubart C, Cuesta J, Felder D (2005) Phylogeography of *Pachygrapsus transversus* (Gibbes, 1850): The effect of the American continent and the Atlantic Ocean as gene flow barriers and recognition of *Pachygrapsus socius* Stimpson, 1871 as a valid species. Nauplius 13: 99–113.

Srivathsan A, Meier R (2012) On the inappropriate use of Kimura-2-parameter (K2P) divergences in the DNA-barcoding literature. Cladistics 28: 190–194. https://doi.org/10.1111/j.1096-0031.2011.00370.x

Taberlet P, Coissac E, Hajibabaei M, Rieseberg LH (2012) Environmental DNA. Molecular Ecology 21: 1789–1793. https://doi.org/10.1111/j.1365-294X.2012.05542.x

Tamura K, Stecher G, Kumar S (2021) MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Molecular Biology and Evolution 38: 3022–3027. https://doi.org/10.1093/molbev/msab120

Thiercelin N (2016) Impact of life history and ecology on rate of diversification and speciation, as exemplified by thoracotreme crabs along the western tropical Atlantic and on both sides of the Isthmus of Panama, 180 pp. https://doi.org/10.5283/epub.32545

Vivien R, Holzmann M, Werner I, Pawlowski J, Lafont M, Ferrari BJD (2017) Cytochrome c oxidase barcodes for aquatic oligochaete identification: Development of a Swiss reference database. PeerJ 5: e4122. https://doi.org/10.7717/peerj.4122

Wolff T (1954) Occurrence of Two East American Species of Crabs in European Waters. Nature 174: 188–189. https://doi.org/10.1038/174188a0

WoRMS (2025) World Register of Marine Species – *Pachygrapsus* Randall. https://www.marinespecies.org/aphia.php?p=taxdetails&id=106965 [June 4, 2025]

Young RG, Abbott CL, Therriault TW, Adamowicz SJ (2017) Barcode-based species delimitation in the marine realm: a test using Hexanauplia (Multicrustacea: Thecostraca and Copepoda). Genome 60: 169–182. https://doi.org/10.1139/gen-2015-0209

Zhang D-X, Hewitt GM (1996) Nuclear integrations: Challenges for mitochondrial DNA markers. Trends in Ecology & Evolution 11: 247–251. https://doi.org/10.1016/0169-5347(96)10031-8

Supplementary material 1

Sampling locations of genetic sequences

Authors: Cátia Alves

Data type: docx

Explanation note: Full dataset of COI 5' region sequences of *Pachygrapsus* species, with correspondence between COI accession numbers and localizations.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.9.152554.suppl1

Supplementary material 2

Phylogenetic tree

Authors: Cátia Alves

Data type: pdf

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.9.152554.suppl2